Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 115 (2017) 274–281

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India

A Novel Web Fraud Detection Technique using Association Rule Mining

Diwakar Tripathi[a,*], Bhawana Nigam[b], Damodar Reddy Edla[a]

[a]*Department of Computer Science and Engineering, NIT Goa, Ponda, Goa, India*
[b]*Department of Information Technology, IET-DAVV Indore, India*

Abstract

In the present scenario online web advertising is premier source of revenue for many internet web applications. Phishing is an activity of misleading web users to fraudulent web sites which can be used to steal the sensitive information from internet. In this article, we introduce a new architecture for web fraud detection using Apriori algorithm for association rule mining and phish tank database in web advertising network. Extensive experiments are done on the proposed architecture with web access log and results obtained by proposed architecture in terms of accuracy, error rate, memory used and search time are encouraging.

Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances in Computing & Communications.

Keywords: Association Rule Mining; Apriori Algorithm; Internet Fraud;

1. Introduction

The term online fraud [1] can be interpreted as social bulling, phishing or steal of sensitive information. In most of the cases online frauds are carried out on naive internet users. In advertisement networks, fraud is carried out by baffling the internet user, false clicks and pop-up ads etc. In this abstraction, the alternation advertisement arrangement is advised and able apprehension alignment is suggested. In such affectionate of advertisement frauds, an alternation of advertisements are produced on the applicant end apparatus (this can be appeared through pop-up or web pages). These advertisements include interactive ads or overwritten URLs that forward a user to the unknown hosts.

Fraud can be termed as mistreatment of profit organization's system which does not results essentially in direct permissible consequences. In today's competitive environment main issue in doing business is threat of fraud, if it's terribly prevailing and if the bar procedures are not secure. This fraud has become important in all the foremost established industry/government data processing applications. Criminal gets space to perform fraud because of huge potential applications of computer networks. Various types of scams or frauds that are performed for years through

* Corresponding author. Tel.: +91-772-007-7622; fax: +0-000-000-0000.
E-mail address: diwakartripathi@nitgoa.ac.in, bhawnanigam@gmail.com, dr.reddy@nitgoa.ac.in

1877-0509
Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances in Computing & Communications
10.1016/j.procs.2017.09.135

mail or telephone call are available and can be seen on the net. Hence fraud detection has become a crucial part of general fraud management. It automates and helps to cut back the manual part of screening or checking method.

There are different types of frauds described [1],[12], in this work we have considered four types of frauds and tried to provide fraud detection architecture. Due to web advertising network, when a user searches free software or some applications at a search engine then it provides the address of web site or sites where the user can get the software like abc.com. The four cases of frauds are listed below.

- Case 1:- abc.com is the address of fraud web site which sends malicious data.
- Case 2:- abc.com addresses to xyz.com, xyz.com addresses to pqr.com, pqr.com addresses to abc.com and hence it forms a loop of sites.
- Case 3:- abc.com addresses to xyz.com, xyz.com addresses to pqr.com, pqr.com displays that it will provide the free software but you have to download another software first which is paid.
- Case 4:- we clicked on abc.com once but multiple pop-up windows are opened against this click, because these sites are generating revenue from advertising vendor on pay per click basis.

Rest of the paper is organized as follows: section 2, presents the background and survey of techniques used in web fraud detection. Section 3, covers the proposed architecture for web fraud detection. Section 4, presents experimental analysis on proposed work. Finally, the paper is concluded based on experimental observations in section 5.

2. Background

This section presents concise survey on various existing fraud detection methods around the internet based advertising fraud networks. Recognition of click fraud in pay per click stream on online advertising network or online is proposed by [2], [3], [4], [5].

Zhang et al. [3] have presented a technique to identify fraud in pay per click model. According to their description as uses of Internet increases, the online advertisement takes essential part in the advertising market. For online advertising, one among the common and wide used revenue models is ”charging for each and every click” based on the certain factors like keywords, popularity and the amount of competing advertisers etc. This pay-per-click model have the result that individuals or opponent companies get the scope to generate the false clicks or can be said as fraud clicks, which causes major problems to the creation of beneficial and strong online advertising market. For the detection of click fraud, an essential issue is to detect the duplicate clicks in jumping windows and sliding windows. These window models can be very effective in defining and determining click fraud. However, many algorithms are available to detect duplicates; there is still a need of practical, realistic and effective solutions to detect click fraud in pay-per-click streams over decaying window architectures.

Haddadi [2] have proposed a model to deal with online click-fraud. They defined bluff ads are group of ads that join forces with the intention of increasing the effort level for click-fraud spammers. These ads can be either targeted ads, with inappropriate display text message, or highly relevant display message, with inappropriate targeting information. Their model works as a check test for legitimacy of the individual clicking on ads. Along with standard threshold-based methods, fake ads help to decrease click-fraud levels.

An advertising network model, using online algorithms has been proposed by [5]. It works on cumulative data to accurately and practically find automated traffic, preserve victim’s privacy, while not changing the business logic model. They propose an absolute classification of the hit inflation techniques and a stream analysis technique that detects a wide range of fraud attacks. They summarized the detection of fraud attacks of some classes as theoretical stream analysis problems that fetch to the data management research community as open problem. A framework has been drawn around for deploying the proposed detection algorithms on a generic model. They bring to a close some successful preliminary findings of their effort to detect fraud on network.

Another model by [6] developed a fraud detection technique for web fraud detection by generating the Streaming-Rules that is based on association between pairs of elements in streams.

Goodman advised a model [7] called pay-per-percentage of impression in which they described an easy technique for mercantilism advertising. Pay-per-percentage of impression deals against each click fraud and impression fraud. A Duplicate Detection in click streams has been proposed by Metwally et al. [[8]]. in this work, they worked they